(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
19 September 2002 (19.09.2002)

PCT

(10) International Publication Number
WO 02/073593 A1

(51) International Patent Classification[7]: G10L 11/00, G11B 20/00, G10H 1/00

(21) International Application Number: PCT/EP02/01719

(22) International Filing Date: 19 February 2002 (19.02.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
01106232.0 14 March 2001 (14.03.2001) EP

(71) Applicant (for all designated States except US): INTERNATIONAL BUSINESS MACHINES CORPORATION [US/US]; New Orchard Road, Armonk, NY 10504 (US).

(71) Applicant (for LU only): IBM DEUTSCHLAND GMBH [DE/DE]; Pascalstrasse 100, 70569 Stuttgart (DE).

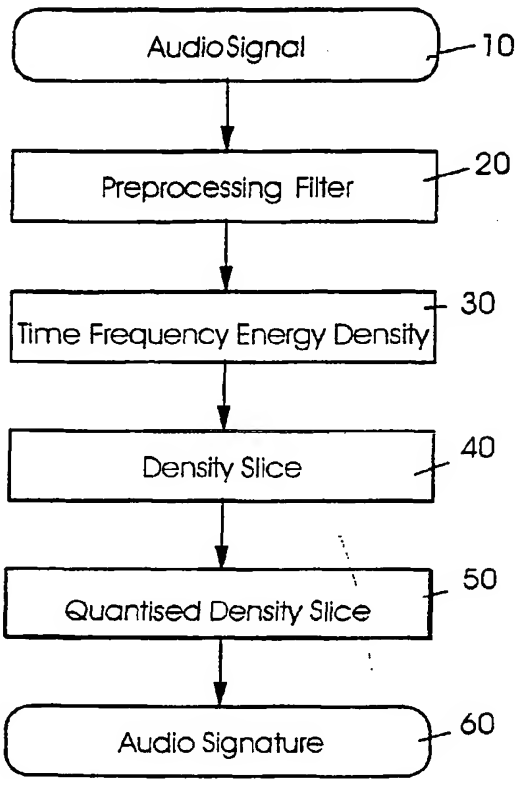
(72) Inventors; and
(75) Inventors/Applicants (for US only): FISCHER, Uwe [DE/DE]; Glashauweg 4, 71088 Holzgerlingen (DE). HOFFMANN, Stefan [DE/DE]; Bäumlesweg 13, 71093 Weil im Schönbuch (DE). KRIECHBAUM, Werner [DE/DE]; Bei der Linde 2, 72119 Ammerbuch-Breitenholz (DE). STENZEL, Gerhard [DE/DE]; Heubergstr. 18, 71083 Herrenberg (DE).

(74) Agent: KAUFFMANN, Wolfgang; IBM Deutschland GmbH, Intellectual Property, Pascalstr. 100, 70548 Stuttgart (DE).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,

[Continued on next page]

(54) Title: A METHOD AND SYSTEM FOR THE AUTOMATIC DETECTION OF SIMILAR OR IDENTICAL SEGMENTS IN AUDIO RECORDINGS



(57) Abstract: Disclosed are a computerized method and system for the identification of identical or similar audio recordings or segments of audio recordings. Identity or similarity between a first audio segment of a first audio stream and at least a second audio segment of an at least second audio stream is determined by digitizing at least the first audio segment and the at least second audio segment of said audio streams, calculating characteristic signatures from at least one local feature of the first audio segment and the at least second audio segment, aligning the at least two characteristic signatures, comparing the at least two aligned characteristic signatures and calculating a distance between the aligned characteristic signatures and determining identity or similarity between the at least two audio segments based on the determined distance.

WO 02/073593 A1

GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) **Designated States** *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

A Method and System for the Automatic Detection of Similar or Identical Segments in Audio Recordings

Field of the Invention

The invention generally relates to the field of digital audio processing and more specifically to a method and system for computerized identification of similar or identical segments in at least two different audio streams.

Background of the Invention

In recent years an ever increasing amount of audio data is recorded, processed, distributed, and archived on digital media using numerous encoding and compression formats like e.g. WAVE, AIFF, MPEG, RealAudio etc. Transcoding or resampling techniques that are used to switch from one encoding format to another almost never produce a recording that is identical to a direct recording in the target format. A similar effect occurs with most compression schemes where changes in the compression factor or other parameters result in a new encoding and a bit-stream that bears little similarity with the original bit-stream. Both effects make it rather difficult to establish the identity of one audio recording and another audio recording, i.e. identity of the two originally produced audio recordings, when the two recordings are stored in two different formats. Establishing possible identity of different audio recordings is therefore a pressing need in audio production, archiving and copyright protection.

During the production of a digital audio recording usually numerous different versions in various encoding formats come

into existence during intermediate processing steps and are distributed over a variety of different computer systems. In most cases these recordings are neither cross-referenced nor tracked in a database and often it has to be established by listening to the recordings whether two versions are identical or not. An automatic procedure thus would greatly ease this task.

A similar problem exists in audio archives that have to deal with material that has been issued in a variety of compilations (like e.g. Jazz or popular songs) or on a variety of carriers (like e.g. the famous recordings of Toscanini with the NBC Symphony orchestra). Often the archive number of the original master of such a recording is not documented and in most cases it can only be decided by listening to the audio recordings whether a track from a compilation is identical to a recording of the same piece on another sound carrier.

In addition, copyright protection is a key issue for the audio industry and becomes even more relevant with the invention of new technology that makes creation and distribution of copies of audio recordings a simple task. While mechanisms to avoid unauthorized copies solve one side of the problem, it is also required to establish processes to detect unauthorized copies of unprotected legacy material. For instance, ripping a CD and distributing the contents of the individual tracks in compressed format to unauthorized consumers is the most common breach of copyright today, there are other copyright infringements that can not be detected by searching for identical audio recordings. One example is the assembly of a "new" piece by cutting segments from existing recordings and stitching them together. To uncover such reuse, a method must

be able to detect not similar recordings but similar segments of recordings without knowing the segment boundaries in advance.

A further form of maybe unauthorized reuse is to quote a characteristic voice or phrase from an audio recording, either unchanged or e.g. transformed in frequency. Finding such transformed subsets is not only important for the detection of potential copyright infringements but also a valuable tool for the musicological analysis of historical and traditional material.

Related Art

Most of the popular techniques currently available to identify audio recordings rely on water-marking (for a recent review of state-of-the-art techniques refer to S. Katzenbeisser and F. Petitcolas eds., Information Hiding: Techniques for steganography and digital water-marking, Boston 2000): They attempt to modify the audio recording by inserting some inaudible information that is resistant against transcoding and therefore are not applicable to material already on the market. Furthermore many of today's audio productions are assembled from a multitude of recordings of individual tracks or voices, often produced at a higher temporal and frequency resolution than the final recording. Using water-marks to identify these intermediate data requires water-marks that do not produce an audible artifact through interference when the tracks are mixed for the final audio stream. Therefore it might be more desirable to identify such material by characteristic features and not by water-marks.

A non-invasive technique for the identification of identical audio recordings uses global features of the power spectrum as a signature for the audio recording. It is hereby referred to European Patent Application No. 00124617.2. Like all global frequency-based techniques this method can not distinguish between permutated recordings of the same material i.e. a scale played upwards leads to the same signature than the same scale played downwards. A further limitation of this and similar global methods is their sensitivity against local changes of the audio data like fade ins or fade outs.

Summary of the Invention

It is therefore an object of the present invention to provide a method and system for improved identification of identical or similar audio recordings or segments of audio recordings.

It is another object to provide such a method and system which allow for the detection of not similar recordings but similar segments of recordings without knowing the segment boundaries in advance.

It is another object to provide such a method and system which allow for an automated detection of identical copies of audio recordings or segments of audio recordings.

It is another object to allow a robust identification of audio material even in the presence of local modifications and distortions.

It is yet another object to enable to establish similarity or identity of one audio stream stored in two different formats, in particular two different compression formats.

The above objects are solved by the features of the independent claims. Advantageous embodiments are subject matter of the subclaims.

The concept underlying the invention is to provide an identification mechanisms based on a time-frequency analysis of the audio material. The identification mechanism computes a characteristic signature from an audio recording and uses this signature to compute a distance between different audio recordings and therewith to select identical recordings.

The invention allows the automated detection of identical copies of audio recordings. This technology can be used to establish automated processes to find potential unauthorized copies and therefore enables a better enforcement of copyrights in the audio industry.

It is emphasized that the proposed mechanism improves current art by using local features instead of global ones.

The invention particularly allows to detect identity or similarity of audio streams or segments thereof even if they are provided in different formats and/or stored on different physical carriers. It thereupon enables to determine whether an audio segment from a compilation is identical to a recording of the same audio piece just on another audio carrier.

Further the method according to the invention can be performed automatically and maybe even transparent for one or more users.

The proposed mechanism for the above reasons allows for an automated detection of identical copies of audio recordings. This technology can be used to establish automated processes to find potential unauthorized copies and therefore enables a better enforcement of copyrights in the audio industry.

Brief Description of the Drawings

In the following, the present invention is described in more detail by way of embodiments from which further features and advantages of the invention become evident, where

Fig. 1 is a schematic block diagram depicting computation of an audio signature according to the invention wherein grey boxes represent optional components;

Fig. 2 is a flow diagram illustrating the steps of preprocessing of a master recording according to the invention;

Fig. 3 is a typical power spectrum of a recording of the Praeludium XIV of J.S. Bach's Wohltemperiertes Klavier where a confusion set for the maximal power contains one element, whereas a confusion set for the second strongest peak contains two elements;

Fig. 4 is a segment of a Gabor Energy Density Slice for a frequency of 497 Hz and a scale 1000 computed for the music piece depicted in Fig. 3;

Fig. 5 is a flow diagram illustrating the steps for quantization of a time-frequency energy density slice according to the invention;

Fig. 6 is a histogram plot of the Gabor Energy Density Slice for the segment with frequency 497 Hz and scale 1000 shown in Fig. 4;

Fig. 7 is a cumulated histogram plot of the Gabor Energy Density Slice for the segment with frequency 497 Hz and scale 1000 shown in Fig. 4;

Fig. 8 raw data of a 497 Hz signature computed for the example of Fig. 4, with unmerged runs for the sample master where start and end are presented in sample units;

Fig. 9 are merged data derived from Fig. 8 for the 497 Hz signature, but for a sample master;

Fig. 10 is a flow diagram illustrating computation of the distance between two audio signatures according to the invention;

Fig. 11 is another flow diagram illustrating computation of a Hausdorff distance, in accordance with the invention;

Fig. 12 is a plot of Hausdorff distance between the 497 Hz Signature of the WAVE master and an MPEG3 compressed version with 8kbit/sec of the same recording, as a function of the shift between the master and the test signature;

Fig. 13 shows a set of ellipses as a typical result of a slicing operation in accordance with the invention;

Fig. 14 shows exemplary templates used for finding those segments in candidate recordings point patterns that are similar or identical to those in the template; and

Fig. 15 shows another set of ellipses for which a template like the one shown in Fig. 14 will match the two segments with the filled ellipses depicted herein.

Detailed Description of the Embodiments

Referring to Fig. 1, prior to the computation of the audio signature 60, analog material has to be digitized by appropriate means.

The audio signature described hereinafter is computed from an audio recording 10 by applying the following steps to the digital audio signal:

Preprocessing Filter

Depending on the type of material and the type of similarity desired, the audio data may be preprocessed 20 by an optional filter. Examples for such filters are the removal of tape noise form analogue recordings, psycho-physical filters to model the processing by the ear and the auditory cortex of a human observer, or a foreground/background separation to single out solo instruments. Those skilled in the art will not fail to realize that some of the possible pre-processing filters are better implemented operating on the time-frequency density than operating on the digital audio signal.

Time Frequency Energy Density

Estimate 30 the time frequency energy density of the audio recording. The time frequency energy density $\rho_x(t,\nu)$ of a signal x is defined by

$$E_x = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \rho_x(t,\nu)dtd\nu$$

i.e. by the feature that the integral of the density over time t and frequency ν equals the energy content of the signal. A variety of methods exist to estimate the time energy density, the most widely known are the power spectrum as derived from a windowed Fourier Transform, and the Wigner-Ville distribution.

Density Slice

One or more density slices are determined 40 by computing the intersection of the energy density with a plane. Whereas any orientation of the density plane with respect to the time, frequency, and energy axes of the energy density generates a valid density slice and may be used to determine a signature, some orientations are preferred and not all orientations yield information useful for the identification of a recording: Any cutting plane that is orthogonal to the time axis contains only the energy density of the recording at a specific time instance. Since the equivalent time in a recording that has been edited by cutting out a piece of the recording is hard to determine, such slices are usually not well-suited to detect the identity of two recordings. A cutting plane perpendicular to the energy axis generates an approximation of the time-frequency evolution of the recording and a cutting plane perpendicular to the frequency axis traces the evolution of a specific frequency over time. For many approximations of the time frequency energy density, density slices orthogonal to the frequency axis can be computed without determining the complete energy density. Both, the orientation perpendicular

to the energy axis and the orientation perpendicular to the frequency axis capture enough information to allow the identification of identical recordings. The actual choice of the orientation depends on the computational costs one is willing to pay for an identification and the desired distortion resistance of the signature.

Quantized Density Slice

The density slice is transformed by applying an appropriate quantization 50. The actual choice of the quantization algorithm depends on the orientation of the slice and the desired accuracy of the signature. Examples for quantization techniques will be given in the detailed description of the embodiments. It should be noted, that the identity transformation of a slice leads to a valid quantization and therefore this step is optional.

Two signatures can be compared by measuring the distance between their optimal alignment. In general, the choice of the metric used depends on the orientation of the quantized density slices with respect to the time, frequency, and energy axis of the energy density. Examples for such distance measures are given in the description of the two embodiments of the invention. A decision rule with a separation value depending on the metric is used to distinguish identical from non-identical recordings.

In the following, two different embodiments will be described in more detail.

1. First Embodiment

The first embodiment describes the application of this invention in the special case of density slices orthogonal to the frequency axis of the energy density distribution and a metric chosen to identify identical recordings. The energy density distribution is derived from the Gabor transform (also known as short time Fourier transform with a Gaussian window) of the signal. The embodiment compares an audio recording with known identity, called "master recording" in the following description, against a set of other audio recordings called "candidate recordings". It identifies all candidates that are subsequences of the original generated by applying fades or cuts to beginning or end of the recording but otherwise assumes that the candidates have not been subjected to transformations like e.g. frequency shifting or time warping.

1.1. Preprocessing of the Master

The master recording is preprocessed to select the slicing planes for the energy density distribution as described in the flowchart depicted in Fig. 2. The power spectrum of the signal is computed 100, the frequency corresponding to the maximum of the power spectrum is selected 110, and the confusion set of the maximum is initialized with this frequency. The energy of the next prominent maxima 120 of the power spectrum is compared 130 with the energy of the maximum and the frequencies of these maxima are added 140 to the confusion set until the ratio between the maximum of the power spectrum and the energy at the location of a secondary peaks drops below a threshold 'thres'. The rational behind the confusion set is that for peaks with almost identical energy values, the ordering of the peaks, and therefore the frequency of the maximum of the power spectrum is likely to be distorted by different encoding or compression algorithms. The value of thresh used by the first embodiment is 1.02. As can be seen

from the confusion set, the master recording used as an example in the description of the first embodiment consist of only the frequency 497 Hz (Fig. 4). As slicing plane(s) for the energy densities, the elements from the confusion set are used, and the values computed during preprocessing are either stored or forwarded to module computing the time-frequency energy density.

1.2. Computation of the Time-Frequency Energy Density

For the master recording and all candidates the time-frequency densities for all elements of the confusion set of the spectral maximum are computed. In the first embodiment a time-frequency density S based on the Gabor transform,

$$S_x(t,\nu;h)=\left|\int_{-\infty}^{+\infty}x(u)h^*(u-t)e^{-2j\pi\nu u}\,du\right|^2$$

i.e. a short-time Fourier transform with the Gaussian window

$$h(z)=e^{-z/2\sigma^2}$$

is used. Since the Gabor transform can be computed for individual frequencies, no explicit slicing operation is necessary and only the energy densities for the frequencies from the confusion set are computed. A segment of the time frequency energy density of the left channel of the example master recording for the frequency of 497 Hz and a scale parameter of 1000 is shown in Fig. 4. The slices of the time-frequency energy density are stored or forwarded to the quantization module.

1.3. Quantization of the Time-Frequency Slice

A time-frequency (TF) energy density slice is quantized as described in the flow chart depicted in Fig. 5. Having read

200 a TF energy slice, the power values are normalized 210 to 1 by dividing them with the maximum of the slice. From the normalized slice a histogram is computed 220 and the histogram is cumulated 230. The bin-width for the histogram used in the first embodiment is 0.01. From the cumulated histogram a cut value is selected by determining 240 the minimal index 'perc' for which the value of the cumulated histogram is greater than a constant cut. The constant cut used in the first embodiment is 0.95. In the normalized slice, all power values greater perc * histogram bin-width are selected 250 and for all runs of such values, the start time, the end time, the sum of the power values and the maximal power of the run is determined 260. Runs that are separated by less than gap sample points are merged, and for the merged runs the start time, the end time, the center time, the mean power and the maximal power are computed. The set of these data constitutes the signature of an audio recording for the frequency of the slicing plane and is stored 270 in a database.

1.4. Comparison of quantized time-frequency slices

The first embodiment uses the Hausdorff distance to compare two signatures. For two finite point sets A and B the Hausdorff distance is defined as

$$H(A,B) = \max(h(A,B), h(B,A))$$

with

$$h(A,B) = \max_{a \in A} \min_{b \in B} \|a - b\|$$

The norm used in the first embodiment is the L1 norm.

To establish the similarity between a master signature and a test signature, the first embodiment computes the Hausdorff distances between the master signature and a set of time-

shifted copies of the test signature, therewith determining the distance of the best alignment between master and test signature. Those skilled in the art will not fail to realize that the flowchart depicted in Fig. 10 for this procedure describes the principle of operation only and that numerous methods have been proposed for implementations needing less operations to compute the alignment between a point set and a translated point-set (see for example D. Huttenlocher et al., Comparing images using the Hausdorff distance, IEEE PAMI, 15, 850-863, 1993). The distance measure used is based on the assumption that the master and the test recording are identical except for minor fade ins and fade outs, to detect more severe editing different metrics and/or different shift vectors have to be used.

Now referring to Fig. 10, in a first step 300 the comparison module reads the signatures for the master and the test recording. A vector of shifts is computed 310, the range of shifts checked by the first embodiment is [-2*d,2*d], where d is the Hausdorff distance between the master and the unshifted test recording. The shift vector is the linear space for this interval with a step-width of 10 msec. For each shift, the Hausdorff distance between master signature and the shifted test signature is computed 320 and stored 340 in the distance vector 'dist'. The distance between master and template is the minimum of 'dist', i.e. the distance of the optimal alignment between master and test signature.

A flow for the computation of the Hausdorff distance is shown in Fig. 11. From both the master signature and the test signature the "center" value is selected and stored in a vector 400. For all elements 410 from the master vector M, the

distance to all elements from the test vector T is computed and stored in a distance vector 420. The maximal element of this distance vector is set 430 the distance 'd1'. In the next step for all elements 440 from the test vector T, the distance to all elements from the master vector M is computed and stored in a distance vector 450. The maximal element of this distance vector is set 460 the distance 'd2'. The Hausdorff distance between the master signature and the test signature is set 470 the maximum of d1 and d2.

The decision whether master and template recording are equal is based on a threshold for the Hausdorff distance. Whenever the distance between master and test is less or equal than the threshold both recordings are considered to be equal, otherwise they are judged to be different. The threshold used in the first embodiment is 500.

2. Second Embodiment

The second embodiment describes the application of this invention in the special case of density slices orthogonal to the power axis of the energy density distribution. The embodiment compares one or more audio recordings ("candidate recording") with a template ("master recording") that contains the motif or phrase to be detected. Typically the template will be a time-interval of a recording processed by similar means than described in this emobidment.

Like in the first embodiment the time-frequency transformation used is the Gabor transform. The time-frequency density of a "candidate recording" is computed using logarithmically spaced frequencies from an appropriate interval, e.g. the frequency range of a piano. This logarithmic scale may be translated in

such a way, that the frequency of the maximum of the energy density corresponds to a value of the scale. The time-frequency energy density such computed is sliced with a plane orthogonal to the energy axis. The result of such a slicing operation is a set of ellipses as the ones illustrated in Fig. 13. These ellipses are characterized by a triplet that consists of the time and frequency coordinate of the intersection of the ellipses major axis and the maximal or integral energy of the density enclosed by the ellipse. Standard techniques like those described in the first embodiment can than be used to find those segments in the candidate recordings point patterns that are similar or identical to those in the template. A template like the one shown in Fig. 14 will match the two segments with filled ellipses in Fig. 15. The third coordinate of the triple can be used as a weighting factor to increase the specificity of the alignment, i.e. by rejecting matches where the confusion sets of the energies of aligned ellipses are different.

It should be noted that ridges (R.Carmona et al, Practical Time-Frequency Analysis, Academic Press New York 1998) can be used as an alternative to ellipses resulting from slicing.

C L A I M S

1. A computerized method to determine identity or similarity between a first audio segment of a first audio stream and at least a second audio segment of an at least second audio stream, comprising the steps of:

digitizing at least the first audio segment and the at least second audio segment of said audio streams;

calculating characteristic signatures from at least one local feature of the first audio segment and the at least second audio segment;

aligning the at least two characteristic signatures;

comparing the at least two aligned characteristic signatures and calculating a distance between the aligned characteristic signatures; and

determining identity or similarity between the at least two audio segments based on the determined distance.

2. Method according to claim 1, wherein the characteristic signatures are represented by an energy density.

3. Method according to claim 2, wherein the energy density is represented by time-frequency energy density.

4. Method according to claim 3, wherein the time-frequency energy density is based on a Gabor transform which is computed for individual frequencies.

5. Method according to any of claims 2 to 4, wherein calculating at least one energy density slice by computing the intersection of the energy density with a plane.

6. Method according to any of the preceding claims, wherein calculating the Haussdorff distance to compare the at least two characteristic signatures.

7. Method according to claim 6, wherein using a threshold for the Haussdorff distance.

8. Method according to any of claims 5 to 7, wherein quantizing the energy density slice.

9. Method according to any of the preceding claims, providing a decision rule with a separation value for determining identity or similarity.

10. A system for determining identity or similarity between a first audio segment of a first audio stream and at least a second audio segment of an at least second audio stream, comprising:

means for digitizing at least the first audio segment and the at least second audio segment of said audio streams;

first processing means for calculating characteristic signatures from at least one local feature of the first audio segment and the at least second audio segment;

second processing means for aligning the at least two characteristic signatures;

third processing means for comparing the at least two aligned characteristic signatures and calculating a distance between the aligned characteristic signatures; and

fourth processing means for determining identity or similarity between the at least two audio segments based on the determined distance.

11. System according to claim 10, further comprising means for computing a time frequency energy density.

12. System according to claim 10 or 11, further comprising means for computing a Gabor transform for individual frequencies.

13. System according to any of claims 10 to 12, further comprising processing means for calculating the Haussdorff distance to compare the at least two characteristic signatures.

14. System according to any of claims 10 to 13, further comprising processing means for quantizing the energy density slice.

15. System according to any of claims 10 to 14, comprising processing means for applying a decision rule with a separation value for determining identity or similarity.

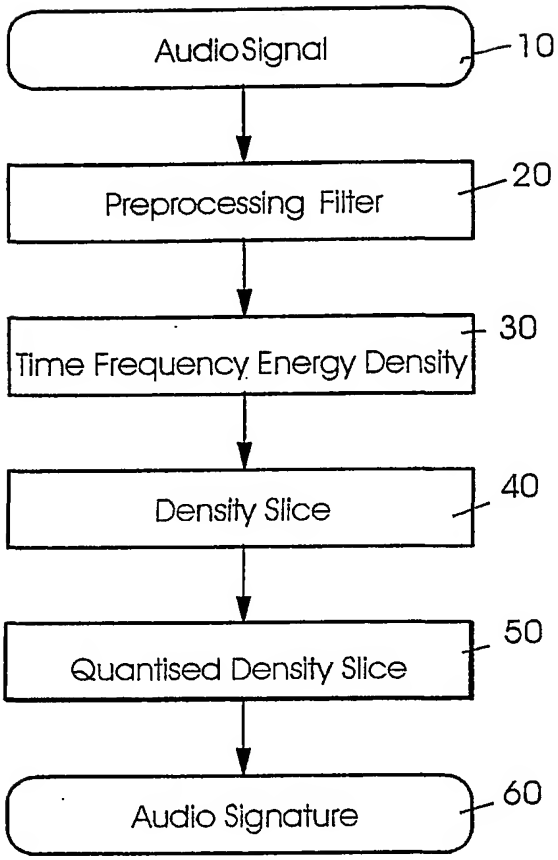

Audio Signal
10
Preprocessing Filter
20
Time Frequency Energy Density
30
Density Slice
40
Quantised Density Slice
50
Audio Signature
60

FIG. 1

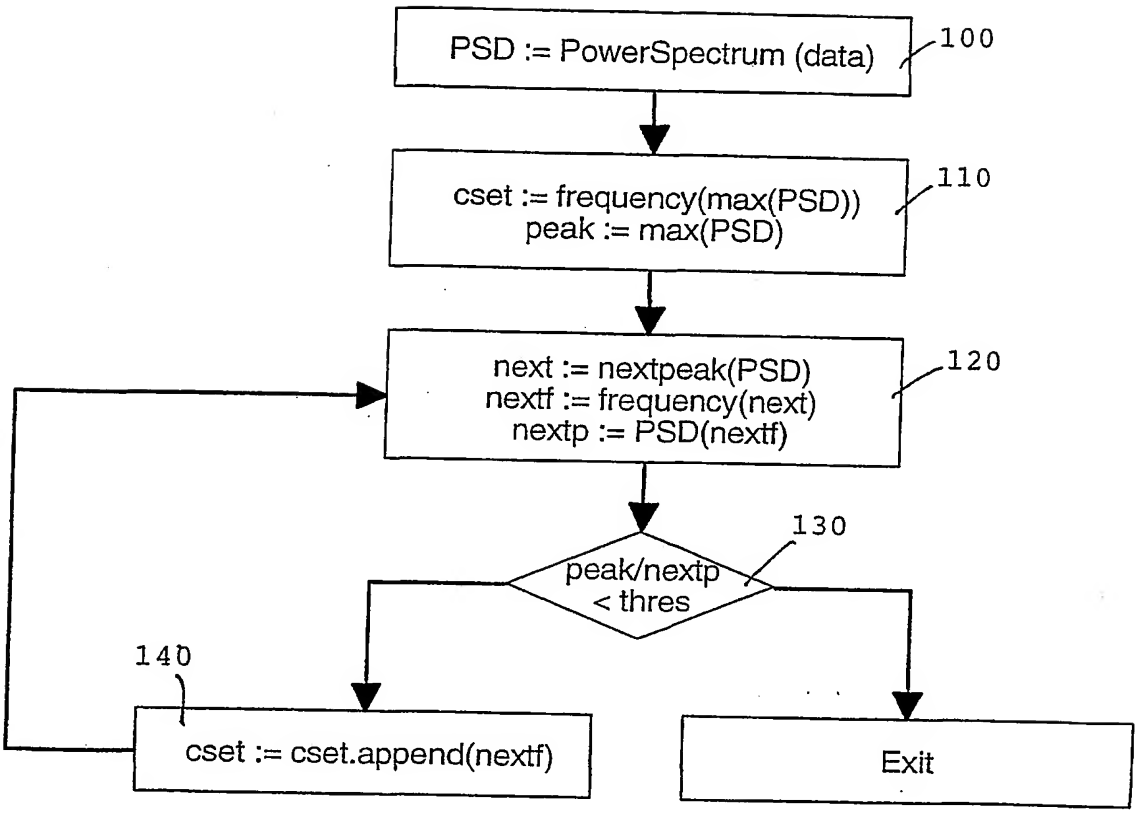

PSD := PowerSpectrum (data)
100
cset := frequency(max(PSD))
peak := max(PSD)
110
next := nextpeak(PSD)
nextf := frequency(next)
nextp := PSD(nextf)
120
peak/nextp
< thres
130
140
cset := cset.append(nextf)
Exit

FIG. 2

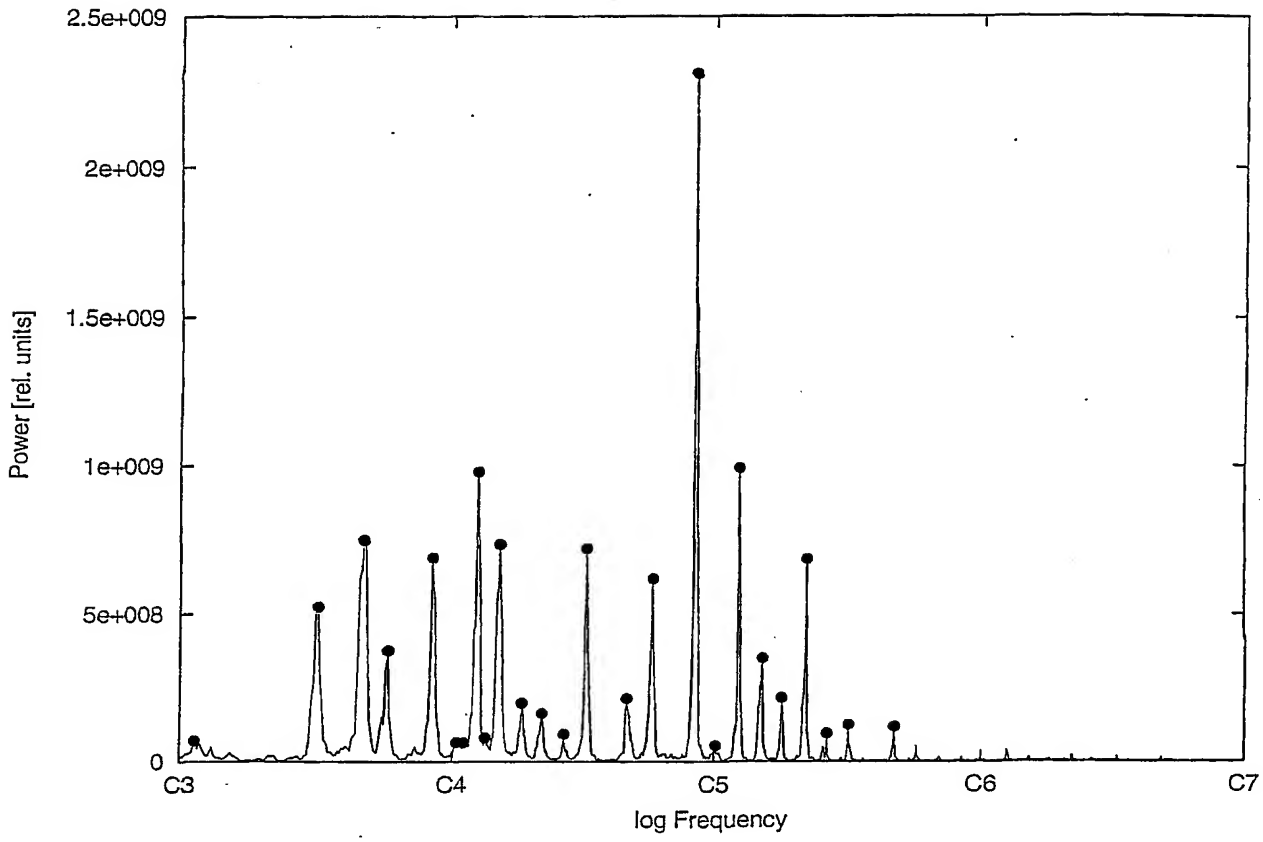

2.5e+009
2e+009
1.5e+009
1e+009
5e+008
0
Power [rel. units]
C3
C4
C5
C6
C7
log Frequency

FIG. 3

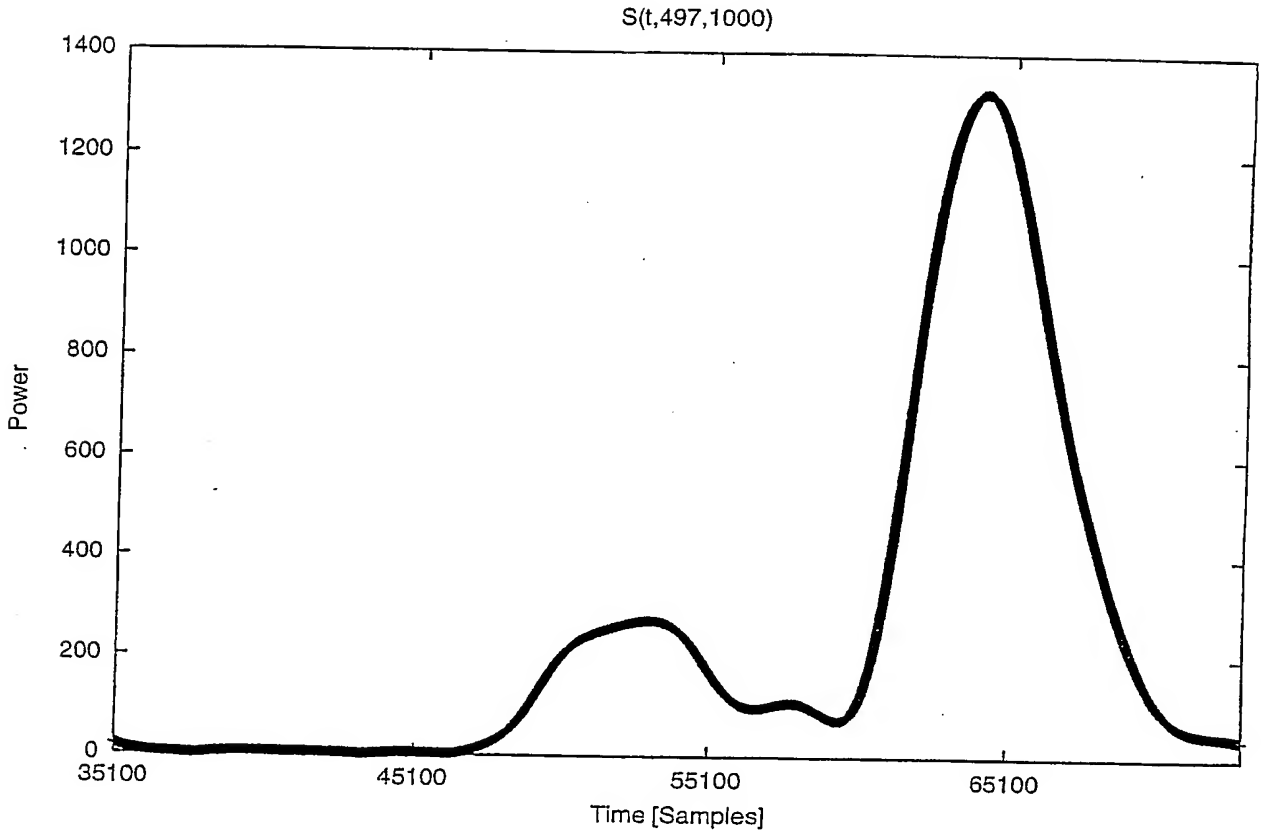

S(t,497,1000)
1400
1200
1000
800
600
400
200
0
Power
35100
45100
55100
65100
Time [Samples]

FIG. 4

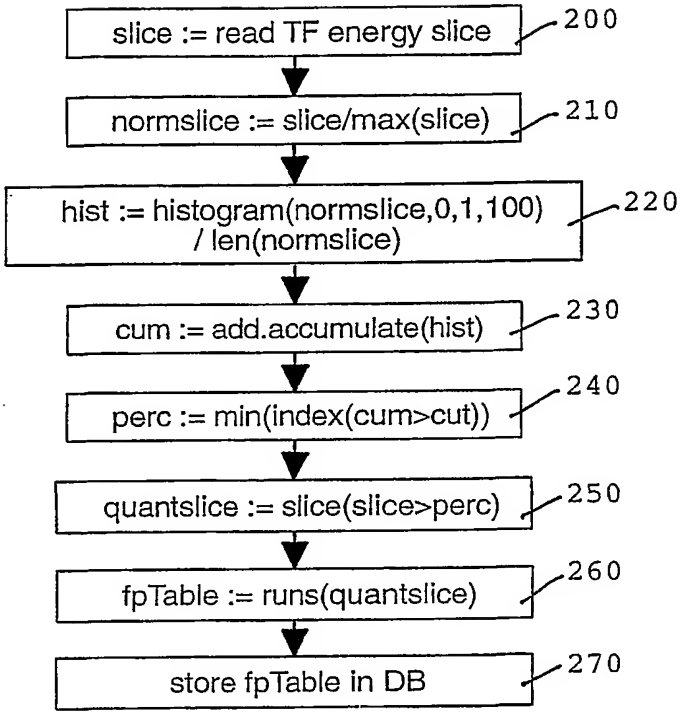

slice := read TF energy slice
200
normslice := slice/max(slice)
210
hist := histogram(normslice,0,1,100) / len(normslice)
220
cum := add.accumulate(hist)
230
perc := min(index(cum>cut))
240
quantslice := slice(slice>perc)
250
fpTable := runs(quantslice)
260
store fpTable in DB
270

FIG. 5

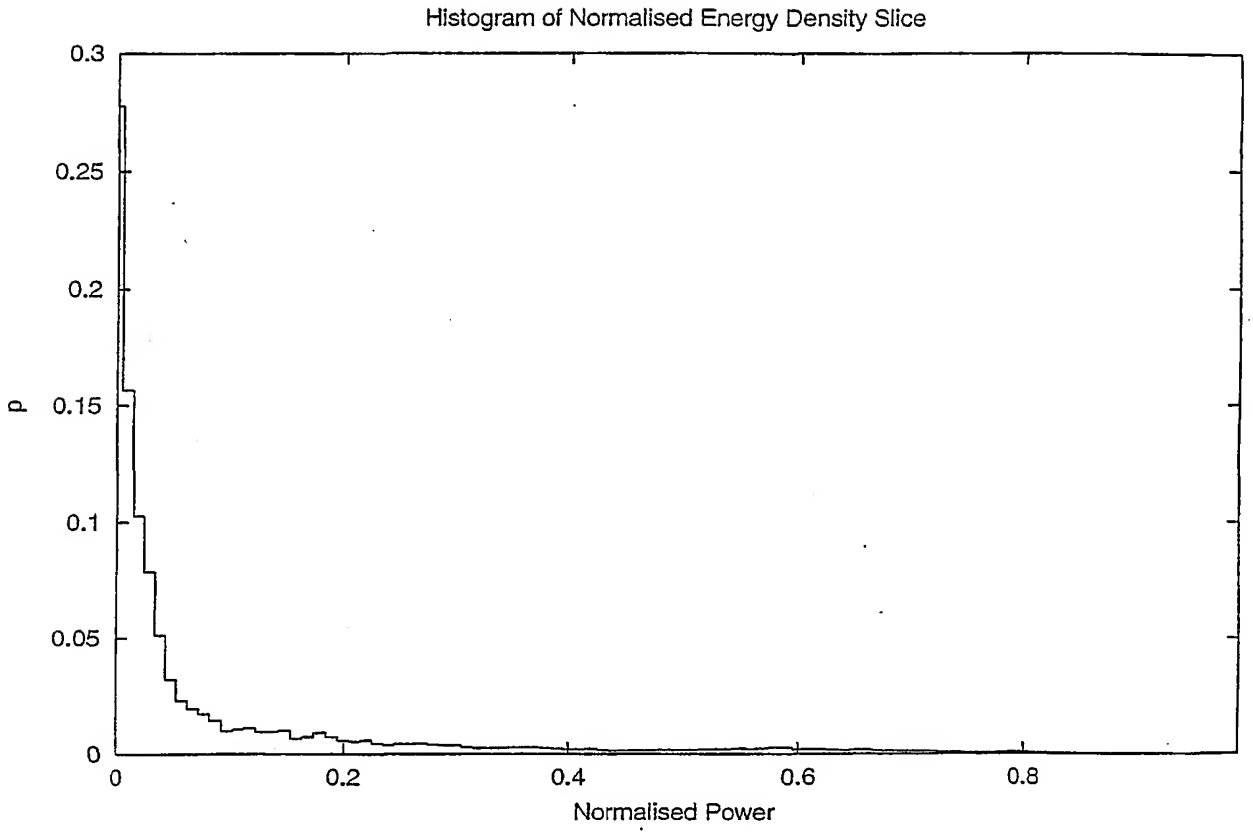

Histogram of Normalised Energy Density Slice
0.3
0.25
0.2
0.15
0.1
0.05
0
p
0
0.2
0.4
0.6
0.8
Normalised Power

FIG. 6

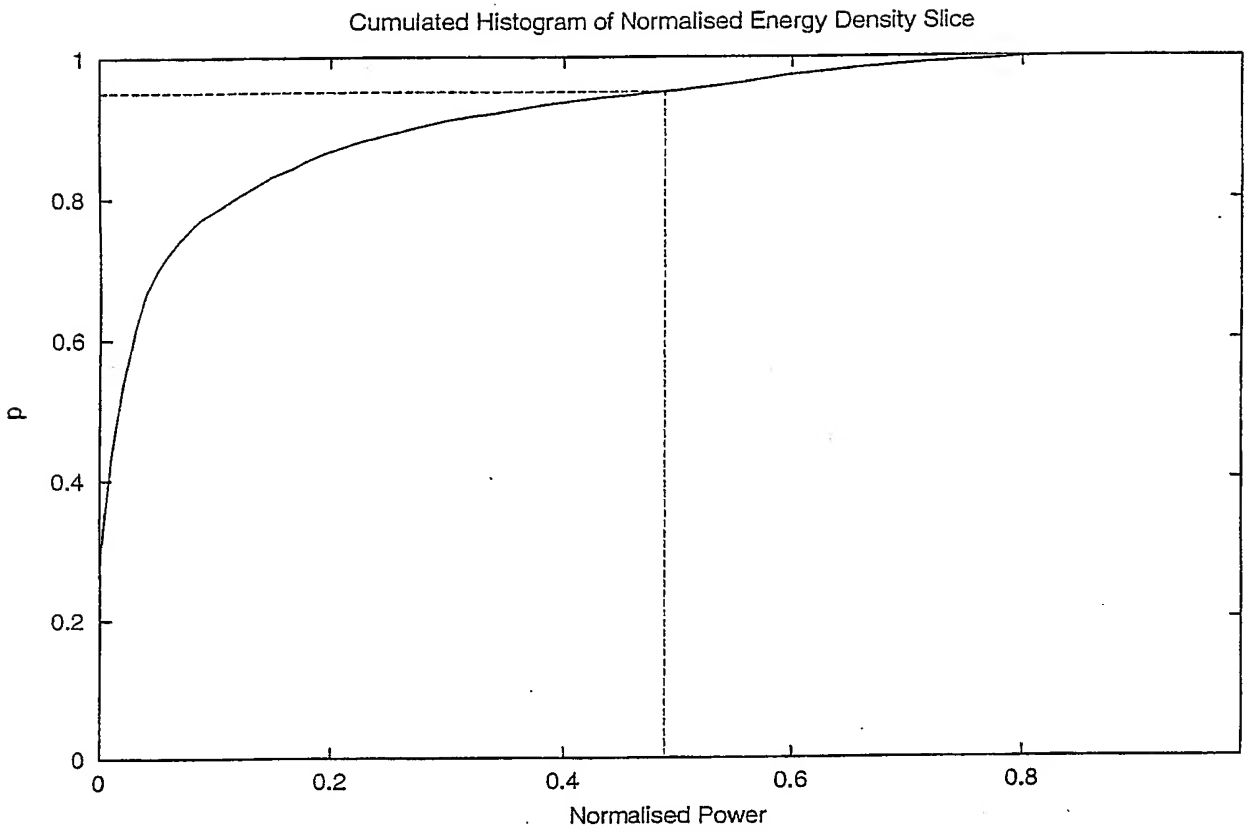

Cumulated Histogram of Normalised Energy Density Slice
p
Normalised Power
0
0.2
0.4
0.6
0.8
1

FIG. 7

Start	Stop	Sum	Max
65092	69207	2672.83	0.73
85082	87220	1120.82	0.54
104787	107205	1298.48	0.56
123063	126450	2038.63	0.66
284321	289893	3173.11	0.62
362817	367456	3225.95	0.81
394290	397629	2019.17	0.67
466047	470608	3156.69	0.81
485100	489102	2355.35	0.64
505223	513495	4850.55	0.63
632794	635232	1348.01	0.59
652742	655096	1267.48	0.56
676747	680445	2349.82	0.71
709083	712433	2022.84	0.66
986375	990599	2642.05	0.67
990600	992950	1258.95	0.58
1524253	1528403	2327.04	0.58
1562485	1566558	2685.77	0.75
1580845	1584953	2835.46	0.81
1600880	1604869	2715.65	0.79
1619492	1624477	3509.47	0.82
1633214	1638299	4419.97	01.0
1638300	1639919	1029.46	0.77
1680094	1683047	1673.52	0.61
1725585	1729999	2826.14	0.72
1818856	1823743	3152.51	0.73
1844088	1849229	3558.33	0.80
1874491	1883032	5898.38	0.82
2034234	2038212	2503.42	0.70
2053761	2057050	1836.99	0.59
2073722	2076589	1645.59	0.62
2092528	2095499	1887.97	0.68
2095500	2096511	565.21	0.62
2107209	2112201	3113.57	0.69
2147372	2150384	1658.69	0.58

FIG. 8

Start	Stop	Certer	Mean	Max
1476.009	1569.319	1522.66	0.64	0.73
1929.297	1977.777	1953.53	0.52	0.54
2376.122	2430.952	2403.53	0.53	0.56
2790.544	2867.346	2828.94	0.60	0.66
6447.188	6573.537	6510.36	0.56	0.62
8227.142	8332.335	8279.73	0.69	0.80
8940.816	9016.530	8978.67	0.60	0.66
10567.959	10671.383	10619.67	0.69	0.80
11000.000	11090.748	11045.37	0.58	0.63
11456.303	11643.877	11550.09	0.58	0.62
14349.070	14404.353	14376.71	0.55	0.58
14801.405	14854.784	14828.09	0.53	0.56
15345.736	15429.591	15387.66	0.63	0.71
16078.979	16154.943	16116.96	0.60	0.66
22366.780	22515.873	22441.32	0.59	0.66
34563.560	34657.664	34610.61	0.56	0.58
35430.498	35522.857	35476.67	0.65	0.75
35846.825	35939.977	35893.40	0.69	0.80
36301.133	36391.587	36346.36	0.68	0.78
36723.174	36836.213	36779.69	0.70	0.82
37034.331	37186.371	37110.35	0.81	1.00
38097.369	38164.331	38130.85	0.56	0.60
39128.911	39229.002	39178.95	0.64	0.71
41243.900	41354.716	41299.30	0.64	0.72
41816.054	41932.630	41874.34	0.69	0.79
42505.464	42699.138	42602.30	0.69	0.82
46127.755	46217.959	46172.85	0.62	0.70
46570.544	46645.124	46607.83	0.55	0.59
47023.174	47088.185	47055.68	0.57	0.61
47449.614	47539.931	47494.77	0.61	0.68
47782.517	47895.714	47839.11	0.62	0.69

FIG. 9

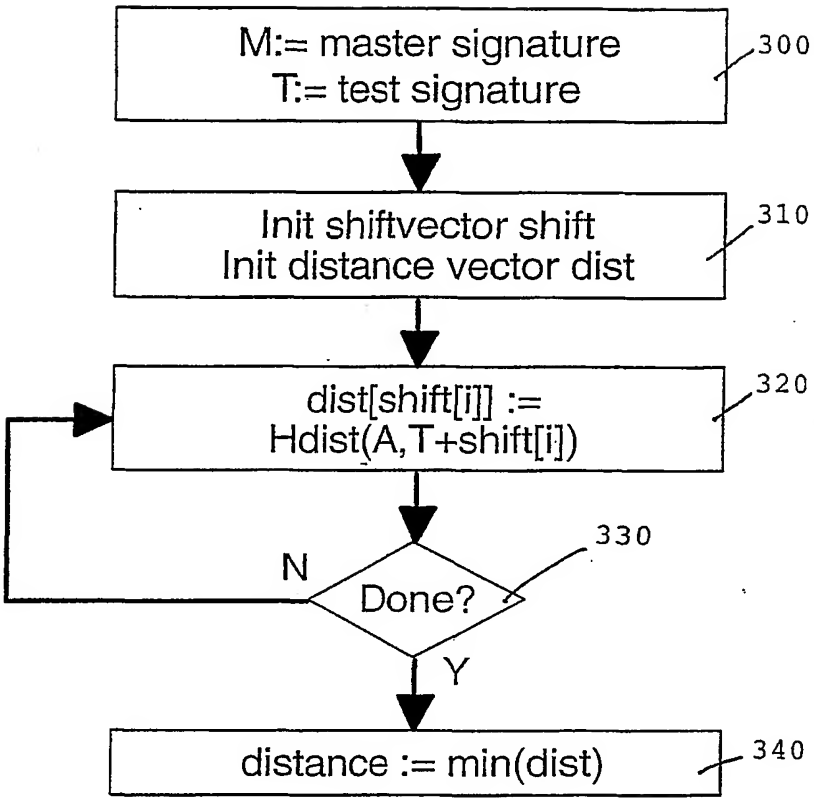

M:= master signature
T:= test signature
300
Init shiftvector shift
Init distance vector dist
310
dist[shift[i]] :=
Hdist(A,T+shift[i])
320
N
Done?
330
Y
distance := min(dist)
340

FIG.10

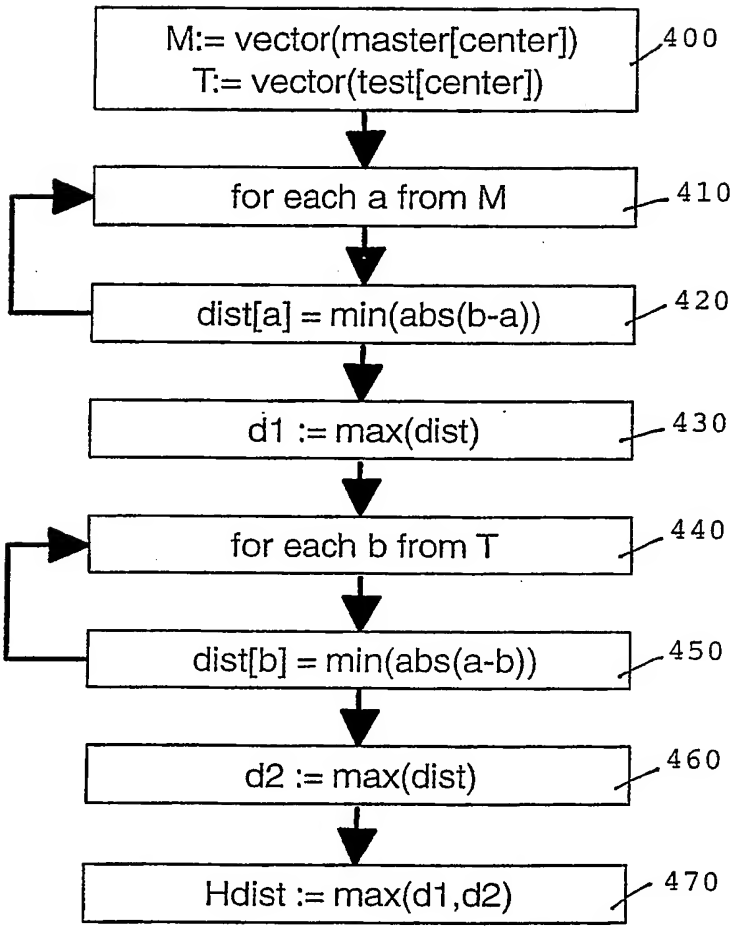


FIG.11

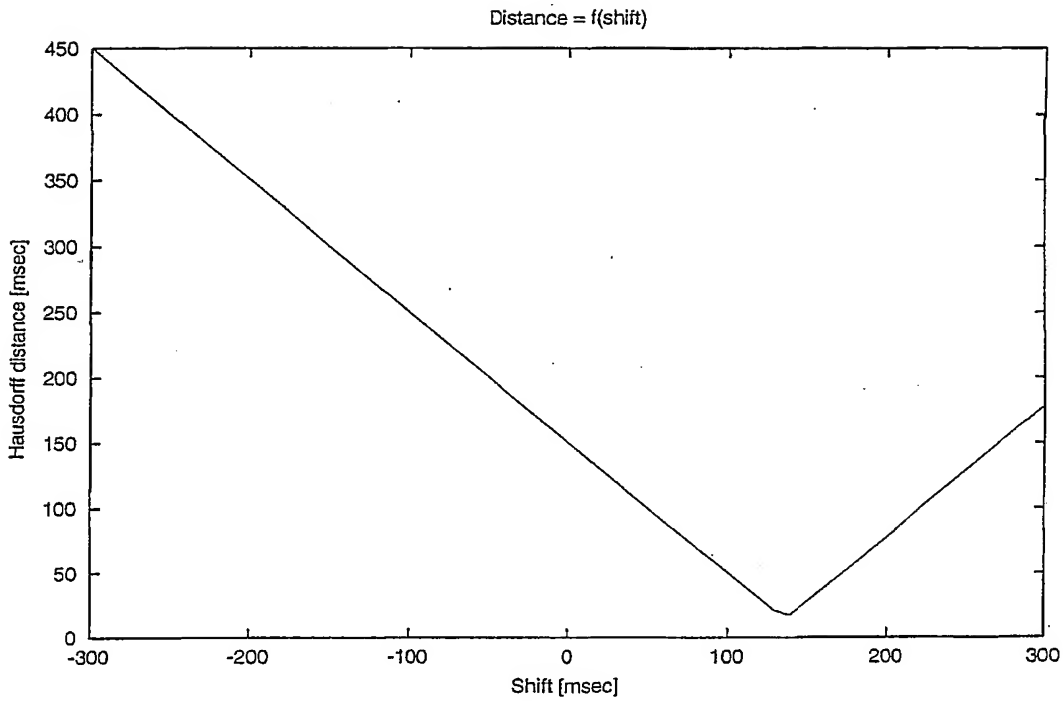

Distance = f(shift)
Hausdorff distance [msec]
450
400
350
300
250
200
150
100
50
0
-300
-200
-100
0
100
200
300
Shift [msec]

FIG.12

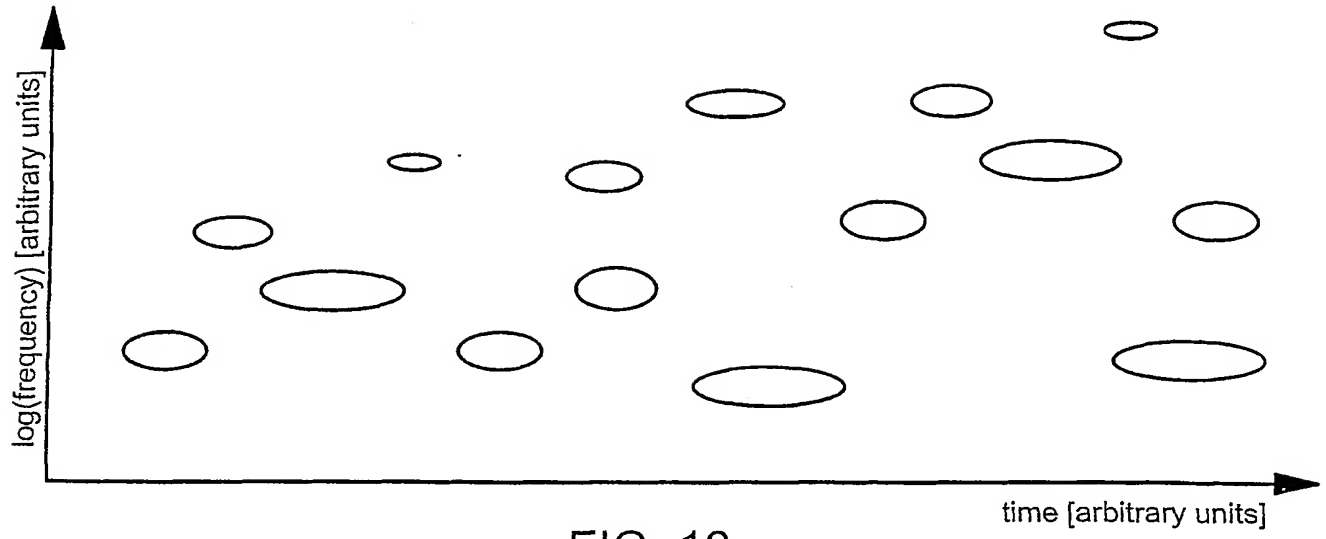


FIG. 13

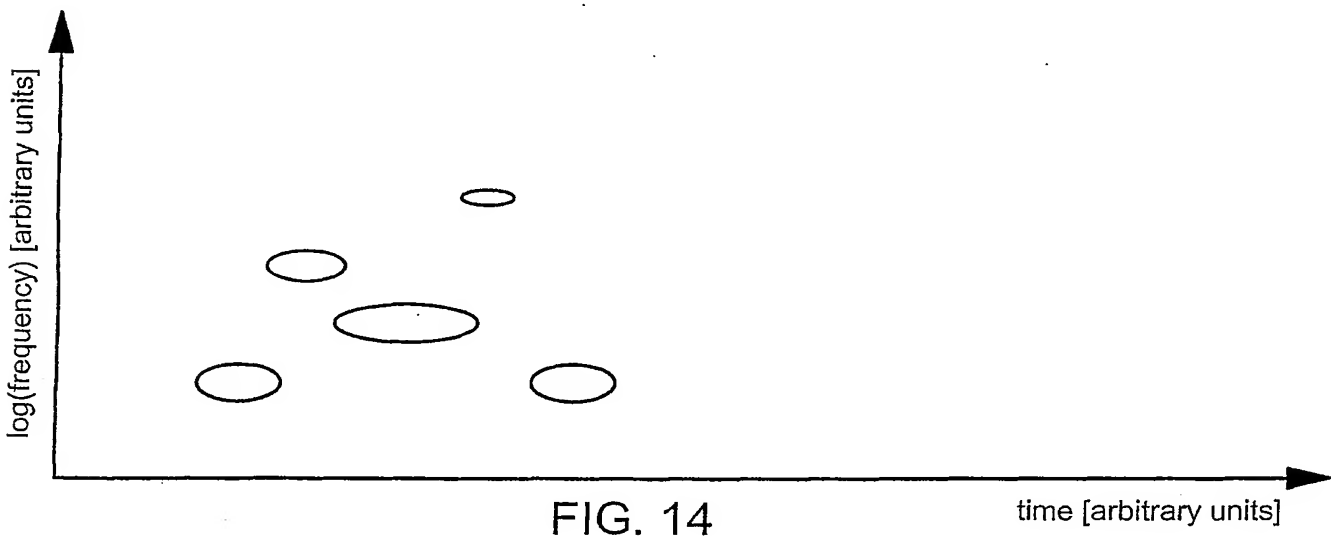


FIG. 14

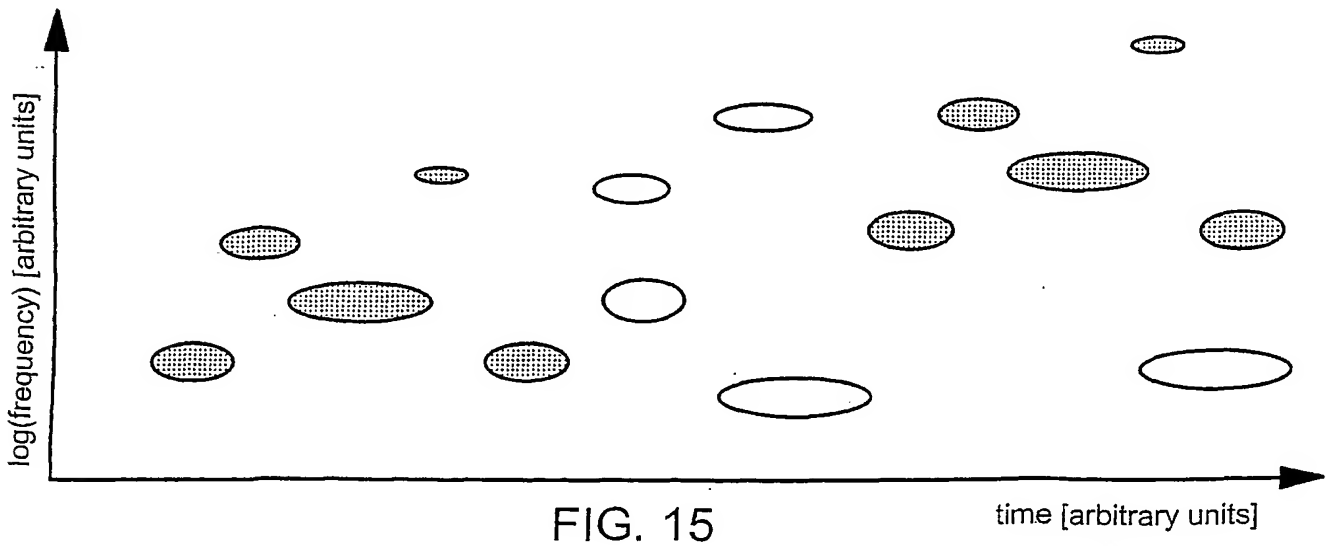


FIG. 15

INTERNATIONAL SEARCH REPORT

International Application No
PCT/EP 02/01719

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 G10L11/00 G11B20/00 G10H1/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 G10L G11B G10H

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, WPI Data, INSPEC

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X Y	US 5 918 223 A (BLUM ET AL) 29 June 1999 (1999-06-29) abstract column 2, line 51 -column 3, line 29 column 3, line 62 -column 4, line 13 ---	1-3, 9-11,15 4,6-8, 12-14
X	WO 01 04870 A (FRAGOULIS DIMITRIOS ;PANAGOPOULOS ATHANASIOS (GR); PAPAODYSSEUS CO) 18 January 2001 (2001-01-18) abstract --- -/--	1,9,10, 15

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
30 May 2002	13/06/2002

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31–70) 340–2040, Tx. 31 651 epo nl, Fax: (+31–70) 340–3016	Quélavoine, R

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

International Application No
PCT/EP 02/01719

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
Y	PAUL D ET AL: "Filterbank implementation of a window based Gabor transform" 1996 CANADIAN CONFERENCE ON ELECTRICAL AND COMPUTER ENGINEERING. CONFERENCE PROCEEDINGS. THEME: GLIMPSE INTO THE 21ST CENTURY (CAT. NO.96TH8157) , vol. 2, 1996, pages 774-777, XP002200570 CALGARY, ALTA., CANADA, New York, NY, USA, IEEE, USA ISBN: 0-7803-3143-5 abstract	4,12
Y	DONG-GYU SIM ET AL: "Pyramidal robust Hausdorff distance for object matching" IMAGE PROCESSING, 1999. ICIP 99. PROCEEDINGS. 1999 INTERNATIONAL CONFERENCE ON KOBE, JAPAN 24-28 OCT. 1999, PISCATAWAY, NJ, USA,IEEE, US, 24 October 1999 (1999-10-24), pages 88-92, XP010368634 ISBN: 0-7803-5467-2 abstract	6,7,13
Y	US 5 210 820 A (KENYON) 11 May 1993 (1993-05-11) abstract	8,14
A	PFEIFFER S ET AL: "AUTOMATIC AUDIO CONTENT ANALYSIS" PROCEEDINGS OF ACM MULTIMEDIA 96. BOSTON, NOV. 18 - 22, 1996, NEW YORK, ACM, US, 18 November 1996 (1996-11-18), pages 21-30, XP000734706 ISBN: 0-89791-871-1 * section 4.1 Music Indexing and Retrieval *	1-15

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/EP 02/01719

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 5918223 A	29-06-1999	NONE	
WO 0104870 A	18-01-2001	GR 99100235 A	30-03-2001
		EP 1147511 A1	24-10-2001
		WO 0104870 A1	18-01-2001
US 5210820 A	11-05-1993	AT 142815 T	15-09-1996
		CA 2041754 A1	03-11-1991
		DE 69122017 D1	17-10-1996
		DE 69122017 T2	10-04-1997
		EP 0480010 A1	15-04-1992
		ES 2091328 T3	01-11-1996
		HK 133697 A	24-10-1997
		JP 5501166 T	04-03-1993
		JP 3130926 B2	31-01-2001
		WO 9117540 A1	14-11-1991